Natural Gradient

Topics in Deep Learning

Objective of the Project

Objective: Application of natural gradient to variational auto-encoders.

Overview of Related Work

Link: New insights and perspectives on the natural gradient method 2014

> “ It has been successfully applied to a variety of problems such as blind source separation, reinforcement learning, and neural network training. ”

Summary: This works gives a brief analysis of the natural gradient method and relates it to the Gauss-Newton Method, and explains the parameter invariance of the natural gradient method.

Link: Topmoumoute online natural gradient algorithm `NIPS 2008`

This paper introduces `TONGA` which shows

> “ much faster convergence in computation time and in number of iteration than SGD ”

- Uses a block-diagonal approximation of gradient covariance matrix (Fisher matrix)
- Performance demonstrated on MNIST and artificial Rectangles dataset.

Link: Natural Neural Networks NIPS 2015

> We introduce Natural Neural Networks, a novel family of algorithms that speed up convergence by adapting their internal representation during training to improve conditioning of the Fisher matrix.

- Experiments using large-scale datasets : CIFAR-10, ILSVRC12
- Achieves the same results as BN

Link: Optimizing Neural Networks with Kronecker-factored Approximate Curvature

`ICML2015`

> We propose an efficient method for approximating natural gradient descent in neural networks which we call Kronecker-factored Approximate
> Curvature (K-FAC). K-FAC is based on an efficiently invertible approximation of a neural networkÃ¢â‚¬â„¢s Fisher information matrix which is nei-
> ther diagonal nor low-rank, and in some cases is completely non-sparse.

How is this work different?

- Couldn't find it in the literature survey.

Proposed Methodology

- Use Natural gradients to train a variational auto-encoder.

Alternate:

- Why do we need ELBO?

Datasets

- MNIST
- CURVES
- FACES

Performance Metrics

- As used in the papers about **Hessian free optimization** and **Optimizing Neural Networks with Kronecker-factored Approximate Curvature** we use reconstruction error.

Update 20/01/16

Goals for Week 2

Fisher information matrix arises as unit perturbation norm on the statistical manifold if KL divergence is used to measure distance.

Goals for this week:

- Try other divergence measures and note the differences in the obtained perturbation norm.
- Summarize TONGA paper, and Natural Neural Networks.